Perceptual grouping boosts visual working memory capacity and reduces effort during retention

Candice C. Morey

Cardiff University

Author Note. Candice C. Morey, School of Psychology, Cardiff University. These data were collected at the Cognitive Neuroscience Suite at the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh. Thanks are due to Amy Andrews and Silvana Mareva for assistance with data collection. Data and analysis scripts are available at https://osf.io/fgwm6/. Address correspondence to Candice C. Morey, School of Psychology, 70 Park Place, Cardiff, CF10 3AT, United Kingdom. Email: MoreyC@cardiff.ac.uk.

Abstract

Consistent, robust boosts to visual working memory capacity are observed when colour-location arrays contain duplicate colours. The prevailing explanation suggests that duplicated colours are encoded as one perceptual group. If so, then we should observe not only higher working memory capacity overall for displays containing duplicates, but specifically an improved ability to remember unique colours from displays including duplicates compared with displays comprising all uniquely coloured items. Furthermore, less effort should be required to retain displays as colour redundancy increases. I recorded gaze position and pupil sizes during a visual change detection task including displays of 4-6 items with either all unique colours, two items with a common colour, or three items with a common colour in samples of young and healthy elderly adults. Increased redundancy was indeed associated with higher estimated working memory capacity, both for tests of duplicates and uniquely-coloured items. Redundancy was also associated with decreased pupil size during retention, especially in young adults. While elderly adults also benefited from colour redundancy, spill-over to unique items was less obvious with low redundancy than in young adults. This experiment confirms previous findings and presents complementary novel evidence linking perceptual grouping via colour redundancy with decreased mental effort.

Keywords: working memory, visual short-term memory, perceptual organization, similarity, eye movement, pupillometry, Bayesian inference

Perceptual grouping boosts visual working memory capacity and reduces effort during retention

It is well known that we explicitly remember remarkably little of the visual detail around us (Bredemeier & Simons, 2012; Simons & Levin, 1997). Yet some elements of the visual scene are rapidly encoded and organized. Structural regularities and features that help to parse the scene into objects affect memory differently than other details (e.g., Alvarez & Oliva, 2009; Woodman, Vecera, & Luck, 2003). Effects of perceptual organization on visual memory are clear: colour redundancy (Quinlan & Cohen, 2012), close proximity and connectedness (Woodman et al., 2003), and background consistency at study and test (Alvarez & Oliva, 2009) all boost memory performance. However, it remains uncertain whether these regularities allow more total information to be maintained, thereby easing apparent memory limits, even though theories predict this regardless of whether they assume visual memory limits are discrete (Adam, Vogel, & Awh, 2017; Rouder et al., 2008) or continuous (e.g., Ma, Husain, & Bays, 2014; van den Berg, Shin, Chou, George, & Ma, 2012). With this study, I aimed to discover whether the presence and strength of perceptual regularities modulates memory for other visual details in the display as expected under the assumption that colour regularities promote perceptual grouping.

There is a consensus that duplicate features, especially salient features like colour, are encoded as a group. Quinlan and Cohen (2012) demonstrated this in a visual recognition memory task. Participants observed a display of several coloured shapes, sometimes with two items sharing the same colour or shape. Performance

was better when the to-be-remembered array contained duplicate colours and when one of the duplicated items was probed. Peterson and Berryhill (2013) showed the same phenomena testing memory for colours-in-location. Morey et al. (C. C. Morey, Cong, Zheng, Price, & R. D. Morey, 2015) revealed that uniquely-coloured items in arrays including duplicate colours also benefit from redundancy. This is what one would expect to observe if the formation of a perceptual group based on feature similarity freed up limited working memory capacity, allowing for the encoding of more detail from the remainder of the display. Indeed, models of immediate memory for visual arrays improve when they account for the perceptual regularities in the scene and for the possibility of grouping discrete items (Brady & Tenenbaum, 2013). Brady and Tenenbaum found that including both possibilities improved the fit of the model to data, which suggests that representing the global properties of the array and applying grouping may reflect distinct processes.

Consistently with the possibility that multiple processes underlie perceptual-grouping benefits to visual memory, Morey et al. (2015) showed that the boost in capacity for uniquely-coloured objects from displays with duplicates *did not* occur while participants' attention was occupied with a backwards counting task, though the boost to performance on tests of the duplicate items remained. The robustness of memory for the redundant colours themselves during attentional disruption agrees well with other research similarly showing that effects of visual-spatial organization are preserved when carrying out concurrent tasks. Rossi-Arnaud, Pieroni, and Baddeley (2006) showed that vertically-symmetrical spatial sequences produced higher spans than non-symmetrical patterns even when participants simultaneously performed an attention-demanding verbal sorting task (Furst &

Hitch, 2000). Even though spatial memory was lower overall under their dual-task conditions, the benefit from organization remained. Altogether, this suggests that benefits arising from perceptual organization cannot be solely dependent on the availability of attention or the flexibility to apply a particular strategy, but perhaps specifically using knowledge about perceptual regularities to recover additional detail benefits from the availability of attention.

Morey et al. (2015) also explored eye movement patterns occurring during the visual recognition task, ruling out the possibility that the similarity-based boost to capacity was due to colour duplicates capturing and holding attention. On the contrary, participants were more likely to fixate unique colours, especially in the few seconds between offset of the stimuli and the appearance of the test item. This suggests that processes were engaged during retention to keep the encoded details activated, and that these processes were focused mainly on attempting to recover the uniquely coloured items. Altogether, the observation that feature redundancy increases capacity, possibly by ensuring that there are fewer high-priority items to be re-activated during retention, leads to the novel prediction that encoding arrays with redundant features should *ease* the burden on participants' limited memories, reducing the mental effort needed to maintain the display's features. Possibly, the notable increase in working memory capacity observed with feature redundancy might be accompanied by a corresponding decrease in pupil dilation, widely understood to reflect the mental effort exerted during a task (Beatty, 1982; Kahneman, 1973; Mathot, 2018).

Confirming that perceptual organizational principles ease cognitive functioning could lead to real-life benefits. For instance, we might use knowledge

about easing memory load via feature redundancy to create visual materials that are easier to digest and comprehend. This could be especially valuable for particular populations known to struggle with certain kinds of memory tasks, such as healthy elderly adults. Performance on visual-spatial memory tasks declines consistently with age (Logie & Maylor, 2009; Swanson, 2017). However, it is not instantly obvious that elderly adults would necessarily benefit from this particular application of feature redundancy. Healthy elderly adults are believed to have special difficulties maintaining associations between features under a variety of circumstances (Brown & Brockmole, 2010; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008), including some evidence that they struggle with object-location binding. The gains observed to visual memory in the presence of redundant features appeared in tests of colour-shape or colour-location associations; these benefits may depend on intact ability to encode these relationships. However, there *is* reason to think that perceptual organization benefits visual memory of older adults regardless of any association memory or binding deficit that older adults may experience. Hamilton, Brown, and Rossi-Arnaud found that symmetry in spatial sequences boosted memory performance in elderly as well as young adults (Hamilton, Brown, & Rossi-Arnaud, 2017). Possibly, the perceptual organization that leads to the capacity boost with colour redundancy likewise occurs in the elderly as well as in the young.

In the present study, I replicated the design of Experiment 1 of Morey et al. (2015) in most particulars, but additionally included trials with as many as three colour duplicates and tested samples of young and elderly adults. Increasing the amount of colour duplicates in some displays affords the possibility of learning whether the benefits to memory capacity continuously increase. The novel goals of

this project include testing whether colour duplicates correspond to a reduction in mental exertion as measured by differences in pupillary responses, and learning whether colour duplicates likewise benefit memory in healthy elderly adults. In addition to establishing whether feature redundancy benefits occur in elderly adults, gaze patterns shall be compared with those observed by Morey et al. (2015) to learn whether the distinct patterns they observed replicate in a new sample of young adults, whether those patterns become more extreme when the amount of colour redundancy increases, and whether elderly adults show comparable gaze patterns. The key patterns Morey et al. observed were a greater likelihood of fixating unique items in displays with duplicates, particularly during retention, and earlier fixations to duplicate items during presentation of the stimuli but earlier fixations toward positions previously occupied by unique items during the retention interval. Prior to beginning data collection, the design and intended analysis plans were pre-registered on Open Science Framework (https://osf.io/9ezjh/). Anonymized data, experiment run files, and analysis scripts may also be found there.

Method

Participants

Forty-five healthy community-dwelling volunteers were recruited from Edinburgh and the surrounding region to participate. Eligible individuals were between the ages of 18 and 35 or 65 and 85 and presented with no evidence of mental illness or dementia. Participants were given a £7 honorarium for completing the one-hour experimental session. One elderly participant did not complete her session due to discomfort with the experimental apparatus, leaving N=44. The young group included 3 males and 18 females, ranging in age from 19 to 33

(*M*=23.48, *SD*=3.79). The elderly group included 9 males and 14 females, ranging in age from 65 to 82 (*M*=71.78, *SD*=5.30).

All participants were in apparently good health, reported normal or corrected-to-normal vision, and successfully completed a brief Ishihara test (1966) to confirm that colour vision was normal. Users of glasses and contact lenses were not discriminated against when recruiting either sample; both samples included individuals with corrected and uncorrected normal vision. Individuals in both groups were highly educated relative to the population at large; mean years in post-secondary education was 4.19 (*SD*=1.47) in the young sample and 4.78 in the elderly sample (*SD*=1.81). A Bayesian *t*-test produced slight evidence *against* the hypothesis that the young and elderly samples differed in years of formal education (BF_{H0}=1.92). Note however, that if there were any difference, the advantage would be with the elderly sample.

The initial recruiting goal was to sample 24 participants per age group, and more participants than needed were scheduled to meet this goal. Recruitment fell short in the young sample because several scheduled participants failed to turn up for their appointments. Data analysis began after the recruitment goal for the elderly sample was met, and the results with respect to any group differences were decisive enough that further data acquisition was deemed unnecessary. Note that all reported analyses make exclusive use of Bayesian inference, so stopping rules were irrelevant.

Apparatus and Stimuli

Data were collected in the Cognitive Neuroscience Suite at the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh. Eye

movements were monitored with an EyeLink 1000 tower mount system. Monocular recording was set to track each participants' right eye, recording 1000 samples per second.

Stimulus presentation and response collection were controlled with custom E-Prime 2.0 software. Stimuli consisted of patterns of 1, 4, 5, or 6 coloured squares. On each trial, each square's location was determined by randomly selecting x and y coordinates from the 270 x 201 pixel region in the centre of the 1024 x 768 pixel monitor, with the restriction that each location must be at least 55 pixels from each other location and from the centre of the screen. Colours were selected randomly from a set of 7 colours chosen so as to nearly equate luminance of each of the stimulus colours with the light grey background (RGB value: 150,150,150) while still creating a discriminable set. An initial set of candidates were chosen using a web-based colour scheme applet (WorkWithColor.com), and then these were honed by measuring their luminance on the laboratory monitor using a photometer and adjusting values until they were within 4.9-5.9 cd/m^2, approximately centred around the measured luminance of the grey background. Colours included in this set are shown in Figure 1, with their RGB values. Colours were randomly selected for use in each trial at run time, with the experimental condition determining whether (and how many) duplicates were allowed.

Procedure and Design

Participants completed a visual change detection task in which three factors were manipulated: 1) the amount of redundancy (0, 2, or 3 items of the same colour), 2) whether a duplicate or unique colour was tested, and 3) set size (1, 4, 5, 6). After indicating consent, participants completed the brief colour-blindness

assessment including 6 items from an Ishihara test (1966). They were then introduced to the change detection task, completing a practice block of 8 trials with feedback on accuracy. Calibration of the EyeLink system commenced after the participant completed the practice trials. Just before the experimental trials began, participants were coached on how to perform articulatory suppression. Participants were instructed to repeat the words "two, ten" aloud at a tempo of 2 words per second from the time the fixation appeared until the probe item appeared. Articulatory suppression was imposed to prevent attempts to verbally recode the colour patterns (though this is not strictly essential; Sense, Morey, Prince, Heathcote, & Morey, 2016). An experimenter attended during the whole session to answer questions, monitor eye movement data acquisition, and ensure compliance with articulation instructions.

Each trial began with the appearance of a fixation "+" for 400 ms, then after a 100-ms pause the study array was presented. After 1200-ms exposure, the colours were removed but the outlines of the squares remained as placeholders for a 3000-ms retention interval. At test, a colour re-appeared in one these positions, which was further highlighted by a surrounding circle. This display remained onscreen until the participant indicated their "same" or "change" response by pressing the corresponding button on a Microsoft Sidewinder gamepad. Figure 1 depicts the stimulus presentation, retention, and cued test events. To allow pupil size to stabilize, the response was followed by a 4000-ms delay before the commencement of the next trial.

Participants completed 192 trials each, divided into two equal blocks. Proportional representation of same versus change trials and tests of unique versus

duplicate colours was attempted to avoid biasing participants toward a particular response. Table 1 shows trial counts per participant and set size by whether the correct response was “same” or “change” and by colour redundancy condition. On change trials, the changed colour could be a new, unstudied colour, or a colour from a different item on the trial; for presentations including duplicates, I systematically varied whether a changed probe colour came from unique items or duplicates (e.g., a tested unique item from an array with a duplicate colour may change to a new colour, to the colour of a different unique item, or to the colour of a duplicate). Because these possibilities varied depending on colour redundancy and set size conditions, it was not possible to perfectly equate the proportion of same and change trials or the likelihood of testing unique items compared with duplicates. Ultimately, this design yielded 55% same and 45% change trials, and no participant noted that either seemed more likely. At least one duplicate was present in the display on 68% of trials, but a duplicate was only tested on 31% of trials. Single-item trials were included to unable modelling of attentional lapse likelihood; fewer single-item than multi-item trials were included because the independent variables could not be manipulated in single-item displays. Slight variance in the number of tests per set size occurred naturally because some types of changes (e.g., swapping the colours of two unique items) were not possible when there were three duplicate colours within a 4-item pattern. Three participants in the young sample completed an earlier version of this paradigm which differed only in that it included 216 trials similarly balanced with respect to proportions of same and change trials, and tests of unique items versus duplicates.

After the first block, participants were given a break of at least one minute and offered a glass of water. When the participant was ready to continue, the EyeLink system was freshly calibrated before the second block of trials commenced.

Analysis Methods

Capacity estimates. The analysis plan follows that of Morey et al. (2015). Using the R package *WMCapacity* to implement the Working Memory Modelling using Bayesian Analysis Techniques tool (R. D. Morey, 2011; R. D. Morey & C.C. Morey, 2011), I estimated *k*s for a variety of models attempting to account for potential effects of age group and level of colour redundancy during presentation and at test. WoMMBAT models working memory capacity as described by Rouder et al. (2008), including parameters estimating individuals' guessing bias (g) and tendency to remain focused on the task (z). I coded colour redundancy at presentation and whether a unique or duplicate colour was tested as a single combined variable to avoid implementing an unbalanced design, but in multiple alternative ways so as to compare hypotheses about which levels of these factors differed. It could be the case that more redundancy at study and testing a duplicate colour both increase k, or it could be the case that only one of these variables matters. I therefore created variable that finely coded 5 levels of colour redundancy: high redundancy (i.e., three instances of one colour) at study, with tests of either a duplicate or unique colour; low redundancy (i.e., two instances of one colour) at study, with tests of either a duplicate or unique colour; and no redundancy at study (or test). I compared the fits of models including this 5-level variable on k with models including simpler implementations collapsing either across level of redundancy (e.g., equating two and three instances of one colour) to make three

levels, coding for only presence of redundancy at study (distinguishing between the finer 3-level coding, the grosser 2-level coding) but disregarding test type, and finally coding only for whether a redundant or unique item was tested.

For hypothesis testing, *k* was the parameter of greatest interest, but *k* values depend in part on allowing *g* and *z* to vary in a reasonable manner. The attention parameter *z* was allowed to vary by individual participant and by age group. It is implausible that the tendency for attention to lapse would be influenced by level of colour redundancy or by whether a redundant or unique colour was tested. Because it was plausible that guessing bias may be influenced by all of these variables, I compared models with only participant and age group on the *g* parameter with models adding various codings of the colour redundancy factor. Allowing *g* to vary by participant, age group, and the most specified coding of colour redundancy produced the best fit, so this combination was used for all possible models of colour redundancy effects on *k*.

Gaze variables. As in Morey et al. (2015), I was interested in understanding whether redundant colours were likely to capture attention during encoding or to be re-attended during retention. I investigated this by comparing the time spent fixating each type of item during the presentation and retention periods. An item was counted as fixated only if gaze fell directly within predefined regions of interest, encompassing a 40 x 40 pixel area (10.58 x 10.58 mm) centred on each presented square and around the central fixation. I also replicated Morey et al.'s analysis of how quickly during presentation and retention duplicate and unique colours were fixated. Additionally, I compared pupil size recovery speed during retention as a function of the amount of redundant colour present across conditions and age

groups. Analyses of gaze data were performed using the R package *BayesFactor* (version 0.9.12-2; Morey & Rouder, 2014; Rouder, Morey, Speckman, & Province, 2012).

Results

Capacity estimates

Table 2 provides the hit and correct rejection rates as a function of colour redundancy, set size, and age group. Trial-level data were entered into WoMMBAT (Morey, 2011; R. Morey & C. Morey, 2011) and models varying in the level of colour redundancy and effects of age on *k* were compared using deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). Lower DIC values indicate superior model fits. Five ways of coding colour redundancy were compared, along with potential effects of the age group factor. Model fit statistics are given in Table 3, with models arranged by how colour redundancy was coded, with more complex models toward the top of the table. The simplest codings specify only probe type (i.e., whether a unique or duplicate colour was probed) and only colour redundancy (*gross* indicates two levels of colour redundancy in the presentation, *fine* indicates three). The more complex specifications code for the interaction between probe type and colour redundancy (with separate models examining gross and fine manners of coding colour redundancy). DIC values tended to decrease as model complexity increased. The best-fitting model included the most complex coding of colour redundancy, in which all three levels of redundancy at presentation were recognized, as well as the different probe types, and an interaction between this factor and age indicated that the colour redundancy effects differed somehow for the young and elderly samples.

Mean *k* values estimated from the best-fitting model are plotted in Figure 2, with the young sample means in the left upper panel, elderly values in the right panel. In both samples, the values order much as would be expected based on previous reports of the colour-sharing bonus (e.g., Morey, et al., 2015; Quinlan & Cohen, 2012; Peterson & Berryhill, 2013). In the young sample, the posterior odds that the *k* for duplicate colour tests with maximum colour redundancy are greater than the other values are at least 217:1. Similarly, posterior odds that the *k* for duplicate colour tests in displays with two repetitions are greater than for unique colour tests in the same condition or in the baseline *All Unique* condition are at least 105:1. As Morey et al. (2015) observed, tests of unique colours in the midst of redundancy share these benefits. With maximum redundancy, posterior odds that *k*s for tests of a unique colour exceeded those for all values in the duplicate-2x or all unique conditions were at least 684:1. In the duplicate-2x conditions, posterior odds that the *k* estimate for unique tests exceeded the baseline *k* were 168:1. This constitutes clear evidence that colour redundancy boosted memory capacity even for unique items from the same display, as well as clear evidence that additional redundancy increased these benefits even more.

How did the *k*s differ in the elderly sample? The inclusion of an interaction between age group, level of colour redundancy, and probe type in the winning model indicates that these variables affected *k* differently for the elderly participants, but lower magnitude of the post-hoc comparisons suggest that differences lie only in the degree of effects observed. Again, duplicate colour tests from the maximum redundancy condition produced the highest *k*s, and odds that these *k*s exceeded all other *k*s were at least 37:1. With maximal colour redundancy,

spill-over of the colour-sharing bonus to unique colours became clear: posterior odds that unique colour tests in the maximally redundant duplicate-3x conditions decidedly exceeded performance on unique colour tests in the minimal redundancy (5999:1) and baseline conditions (>25k:1) were convincingly high. However, with less redundancy, posterior odds that *k* for unique colours exceeded the baseline *k* were only about 22:1, which is more tentative evidence than that obtained in the same analysis with young adults. While not drastically different from the pattern observed in young adults, the interaction with age group present in the best-fitting model seems to arise because elderly adults benefited less from low redundancy than young adults did, particularly for unique colour tests. With maximal colour redundancy however, clear benefits emerged in older adults even for unique colour tests. Across all conditions, mean *k*s in the young adult group were 3.87 compared to 3.07 in the elderly group, replicating the age-related decreases to visual memory observed previously (e.g., Logie & Maylor, 2009; Swanson, 2017).

Gaze towards unique and duplicate colours

Five participants (four from the elderly group) were excluded from all analyses of gaze due to poor acquisition of gaze data or inability to calibrate the eye tracker, leaving *N*=39. Table 4 gives the average number of fixation counts and the average time of fixations toward each a priori category of interest areas (centre, unique colour, duplicate colour) for the stimulus presentation and retention periods and the colour redundancy conditions. There was no evidence of differences between the amount of gaze data acquired amongst different colour redundancy conditions or age groups during stimulus presentation (*BFs* from 0.01 – 1.48), but

slight evidence that participants in both age groups committed fewer fixations in conditions with more colour redundancy during the retention period (BF = 3.66).

Under these unrestricted viewing conditions of displays where all items could be seen when fixating the centre, we observed little effect of directly fixating the to-be-tested item on accuracy, which is consistent with previous findings. Evidence suggested that directly fixating the eventual probe item did not influence accuracy (BFs_{Null} > 4.35).

Consistently with Morey et al. (2015), we observed little reason to believe that duplicates captured and held attention. I compared the proportion of time spent fixating duplicates and unique colours in displays with redundancy to what would be expected if fixations to objects were distributed randomly, carrying out a Bayes Factor ANOVA on differences between actual and expected proportional gaze time. This analysis included fixed factors of colour redundancy condition (two or three duplicates), age group (young and old), and set size (4, 5, or 6) on relative looking toward duplicate colours. During stimulus presentation, the best model included main effects of set size and colour redundancy condition (BF > 3900, ± 0.92%), with participants looking at duplicates *less* frequently than expected at set sizes 4 (M= -0.06) and 5 (M= -0.03), and slightly more than expected at set size 6 (M=0.04), and showing a stronger tendency to look toward unique colours when there were three duplicates (M= -0.04) than when there were two duplicates (M=0.01). This pattern did not appear to differ with age: excluding a main effect of age on relative looking was favoured by a factor of BF > 4, and excluding an interaction between age and colour redundancy condition was favoured by BF > 17. Relative time looking at duplicates is plotted in Figure 3 by colour redundancy

condition and set size for both the stimulus presentation and retention periods, along with *BF*s comparing each estimate with chance expectations. With only one estimate numerically greater than chance and some estimates decisively lower than chance, the data in Figure 3 provide no reason to believe that duplicate colours generally captured and held attention during stimulus presentation.

During retention, it became even clearer that participants were more likely to seek out the locations where unique colours had been presented in arrays with duplicate colours. While the best model included only main effects of set size and colour redundancy condition (*BF*> 50 million), a model also including an interaction between set size and age group could not be ruled out (*BF*= 2.10 favouring the simpler model). If there is truly an interaction to be observed here, it appears to be in how strong the tendency to look toward unique colours was, not whether unique colours were more likely to be fixated than duplicates. Both young and old participants consistently fixated the locations that had previously held duplicate colours much less frequently than would be expected by chance when there were two repetitions (*M*= -0.07) and when there were three repetitions (*M*= -0.16). The tendency to seek out the locations of the unique colours appeared to be strongest at set size four in both young (*M*= -0.16) and old (*M*= - 0.17), but dropped nearly to chance level in the older sample as set size increased to 6 (M_{SS5}= -0.10, M_{SS6}= -0.02) while remaining well below chance in the younger sample (M_{SS5}= -0.14, M_{SS6}= -0.11). Altogether, this replicates Morey et al.'s (2015) finding that during retention individuals tend to look back to the locations of the unique items, and extends it by showing that this trend increases with increased redundancy.

Speed of first fixating unique vs. duplicate colours

In trials containing a duplicate colour, another way to assess attentional capture is to consider which kind of object was fixated first. Morey et al. (2015) showed that during presentation of the stimulus participants were likely to fixate a duplicate earlier than a unique colour, but during the retention period, they were more likely to quickly fixate the position occupied by a unique colour, in addition to looking more frequently at the unique-colour positions overall. This pattern suggests that duplicates may capture attention initially, but that later participants may focus on recalling the unique colours. This explanation is consistent with the proportional fixation patterns, which suggest that during retention former positions of unique colours are fixated decisively longer than duplicates, and suggest a schema for detailed encoding based on quickly-acquired knowledge about the array (namely, how it is structured with respect to the duplicate colours). However, since in this sample eventual trial accuracy did not predict whether the participant had directly fixated the probed item, these results should be interpreted cautiously.

I replicated Morey et al.'s (2015) analysis in this sample. I analysed the minimum time taken per trial to fixate three potential interest areas (the centre of the screen, the position of a unique colour, or the position of a duplicate) during the stimulus presentation and retention periods. For each period, I began with an omnibus analysis including age group (elderly or young), colour redundancy condition (two or three duplicates), and set size (4, 5, or 6) in addition to interest area. I used the omnibus analysis to rule out factors, justifying simpler follow-up analyses. During stimulus presentation, set size was not present in the competing models with the highest Bayes factors; exclusion of set size from the best model was favoured by a factor of at least 60. Follow-up analyses including the remaining

factors supported a model including effects of interest area, age group, and an interaction between interest area and age group (BF=1.86x10^{690}, ± 1.77%). Inclusion of the interaction between age group and interest area was strongly favoured (BF>1 million). Though older adults' minimum first fixations (M=905, SD=528) were slower overall than those of younger adults (M=828, SD=537), their fixations off the centre were quicker than those of young adults (see Table 5 and Figure 4). However, in both groups, the centre was likely to be fixated earliest, then a duplicate, and then a unique colour. Excluding colour redundancy from the model was favoured by a factor of more than 30, so this pattern is likely unaffected by the amount of redundancy in the display.

During retention, the opposite pattern with respect to interest areas emerged (similarly to Morey et al., 2015): the positions where unique colours had been presented were likely to be fixated earliest. The omnibus analysis suggested that colour redundancy would be excluded, so I ran a 3-way Bayes factor ANOVA on minimum fixation times including the remaining factors. The best model included only an effect of interest area (BF=1.98x10^{45}, ± 1.86%). An additional effect of set size could not be ruled out (BF for excluding it was 1.85). Excluding an effect of age group was favoured by a factor of more than 10. These results are depicted in Figure 5, which makes clear that speed of fixating the former positions of unique colours during retention was quickest regardless of age group or set size.

Pupillometry

Changes in pupil size can be taken as reflections of the cognitive effort expended during a task (Kahneman, 1973). Though differences in tendencies to fixate unique items could not explain observed differences in capacity estimates

between young and elderly adults, possibly the similar processes occurring in the young and elderly samples required different levels of effort, which could be distinguished via pupillometry. Figure 6, which depicts proportional pupil size (against average baseline area values recorded per participant during the second before the plotted period) beginning 1.50 seconds prior to stimulus presentation in the elderly and young samples as a function of colour redundancy condition, suggests that perhaps the age groups differed in effort expended during retention as a function of the amount of colour redundancy in the to-be-remembered display. A sharp decrease in pupil size was apparent after the onset of the stimulus presentation, indicative of the brightness change in the monitor when the stimuli appeared (which is to be expected; Mathôt, 2018). This makes analysis of the stimulus presentation period impossible, but the rates of pupil size recovery during the 3000-ms retention period can be taken to reflect expended effort.

I entered proportional pupil sizes into a Bayesian mixed linear model including age group, set size, colour redundancy, and time during the retention period (divided into 10 300-ms bins to balance the competing need for fine granularity of the data with ease of computation) as factors. The best model included interactions between age group and each of the other three factors, as well as an interaction between set size and colour redundancy (BF=2.53x10^{812}, ±2.85%). Means (and standard deviations) for both age groups along the set size and colour redundancy values are given in Table 6. The interaction between set size and colour redundancy was favoured by a factor of more than 10,000, and appeared to be due to larger differences between colour redundancy conditions at low compared to higher set sizes. As for the interaction with age group, for elderly adults little

difference emerged in pupil size across set size or colour redundancy conditions, whereas for the young adults, pupil size seemed to gradually increase with set size, and to gradually increase with the number of unique colours to be remembered (see Figure 6). Though the absolute values are small, interactions with age group were favoured decisively: the interactions between age group and colour redundancy and between age group and set size were both favoured more than 5900:1, and the interaction between age group and time period was favoured by many millions. Figure 6 reveals a much larger and more graded difference between the three levels of colour redundancy in the young than in the elderly adults, with more repetition corresponding to smaller pupil values that take longer to increase to baseline during the retention interval. Smaller values in conditions with more redundancy are consistent with the suggestion that encoding duplicate colours requires less effort, consistently with the hypothesis that they are encoded together as group very early in the trial. When more items from the to-be-remembered display are grouped into a single configuration, young adults require less effort to remember the remaining contents of the display.

Discussion

It is now firmly established that feature redundancy in to-be-remembered visual displays drastically boosts memory (Morey et al., 2015; Peterson & Berryhill, 2013; Quinlan & Cohen, 2012). Here, this benefit was confirmed, and novel tests provide additional detail about how and why this benefit occurs. First, Morey et al.'s finding of a measurable carry-over benefit to unique items from displays with redundancy was confirmed. The present experiment revealed continuous boosts to uniquely coloured items across three levels of feature redundancy. Second, I

confirmed that increased feature redundancy is associated with reduced pupil dilation during retention, which suggests that redundancy eases the effort needed for maintenance. Healthy older adults' memories also benefitted from feature redundancy. Though elderly adults interacted with the stimuli in a similar manner as younger adults, they showed less differentiation of the pupillary response with redundancy, suggesting that they exert more effort to perform the same task. This finding is consistent with compensation-based theories of cognitive aging (Park & Reuter-Lorenz, 2008; Reuter-Lorenz & Cappell, 2008), and since there was also no evidence of differences in gaze patterns between age groups which might have reflected the deployment of categorically different strategies, the pupillometry evidence suggests that increased effort in the elderly is unlikely to be strategy-dependent. The differences we clearly observed between young and elderly adults – attenuated easing of effort needed during retention to maintain the arrays, and less clear boosts to memory for unique items amidst redundancy - are consistent with the idea that cognitive resources might be more limited in the elderly. However, these results also confirm that perceptual organization benefits the elderly as well as the young, and does not depend on attention. Altogether, these findings confirm that colour redundancy could be exploited to boost memory for visual materials in both young and older adults.

It should be noted that though memory performance is expressed in terms of *k*, this discrete model of memory capacity was not used in order to attempt to argue that feature redundancy literally expands the capacity of some latent memory store. The hierarchical Bayesian model of *k* applied here (R.D. Morey, 2011; R.D. Morey & C.C. Morey, 2011) is useful for detecting which manipulated factors affect memory

performance. That estimated *k*s increased as feature redundancy increased suggests that feature redundancy affords the possibility for chunking, or that encoding of the perceptual gist of the array allows for more efficient encoding of the items within that structure. That feature redundancy should ease memory in this way is an expectation of visual memory models regardless of whether they assume discrete or continuous limits. At this point, a model of visual memory capacity that posits objective limits should account for encoding the perceptual gist of the display and for allowing that grouping of discrete items may also occur, as Brady and Tenenbaum (2013) demonstrated.

Morey et al.'s (2015) observations about speeded looking toward the unique items during retention replicated in both the young and elderly samples reported here. This pattern, particularly the consistency with which participants apparently seek out the former positions of uniquely-coloured items during retention, is interesting because it eliminates the possibility that participants' gaze (and possibly their attention) during retention is drawn to the items they definitely remember. Participants are much more likely to remember the duplicate items from an array; if their attention during retention were focused on the items held in mind, then we should have observed quicker and more frequent gazes toward the position of duplicate, not unique, colours. The pattern replicated here is consistent with the idea that fixating during retention reflects a covert retrieval attempt (see Ferreira, Apel, & Henderson, 2008). However, in this study no clear benefits of fixating during retention on performance were observed. If re-fixating the position of a to-be-remembered item reflected an attempt to retrieve details about that item, then one would predict that accuracy would be higher for probed items that were fixated

than probed items that were not. Possibly, this process is not uniformly effective, and thus not guaranteed to benefit performance much. The consistency of these patterns across samples suggests that they reflect some stable process. Further experimentation is needed to work out what processes these consistent gaze patterns reflect. Notably, the early looking to redundant items during presentation that Morey et al. documented also replicated. This finding is consistent with the supposition that perceptual organization occurs rapidly and automatically (see discussion of Hamilton et al., 2018). It is also notable that similar organization does not seem to occur for less salient shape features (Quinlan & Cohen, 2012) or for semantic aspects of shapes (Quinlan & Cohen, 2016). Altogether, the available evidence strongly suggests that the apprehension of colour redundancy rapidly organizes the scene, and this organization clearly benefits memory.

The effects of feature redundancies and similarities seem to differ for visual and auditory-verbal materials. Because short-term memory phenomena for serial verbal materials have been studied so extensively, a natural default assumption for visual materials is to assume that the same phenomena would appear in visual memory (Morey, 2018). This leads to the expectation, for instance, that visually similar items will be confused with each other and their presence in a stimulus will lead to increased errors, as with acoustically similar stimuli (Baddeley, 1966; Conrad, 1964). In fact, the evidence for visual similarity effects has been mixed. Visual confusion errors can occur (Awh, Barton, & Vogel, 2007), particularly in the recall of visually-presented verbal information (Logie, Della Sala, Wynn, & Baddeley, 2000; Logie, Saito, Morita, Varma, & Norris, 2016; Saito, Logie, Morita, & Law, 2008) or when visual information must recalled in serial order (Jalbert,

Saint-Aubin, & Tremblay, 2008). Other evidence contrarily indicates that feature similarities do not impair visual memory, and under some circumstances similarity may even enhance visual memory (Jiang, Lee, Asaad, & Remington, 2016; Lin & Luck, 2009; Sun et al., 2017). Sun et al. (2017) found that the presentation of more visually *dissimilar* to-be-remembered items provoked greater interference, increasing the likelihood of wholesale forgetting of visual objects. A boundary condition could be whether the information has a temporal element. In comparing effects of colour similarity on memory for sequences of colours, Jalbert et al. found that recall of sequences of similarly-coloured items were impaired regardless of whether the colours were to be recalled in their correct order or whether they were to be placed in their correct locations. However, while order errors were likely to fall close in temporal space to the correct position, spatial location errors were unrelated to the nearness of the alternatives. The available evidence suggests that similarity may operate differently when recalling details across time versus space. Clearly, more work is needed to fully reveal the boundary conditions on effects of similarity on visual memory, but it already appears that similarity affects visual memories differently from acoustic or verbal memories, perhaps largely because acoustic and verbal memories are likely to include temporal information. Though differences between visual and verbal short-term memory phenomena are often taken as evidence for distinct short-term memory systems (e.g., Quinlan & Cohen, 2016), it is plausible that differences between the boundary conditions of acoustic and visual similarity effects arise due to the differences in perceptual and motor affordances natural to verbal and visual stimuli (Macken, Taylor, & Jones, 2015), as well as the task characteristics (e.g., recall or recognition, sequential or

simultaneous presentation) that frequently differ between verbal and visual measures of ostensibly analogous phenomena (e.g., Ward, Avons, & Melling, 2005).

In conclusion, these results bolster the view that redundant visual features in a display are encoded as a perceptual group, and that this organization facilitates memory both for the grouped items and for uniquely coloured items from the same display. These robust boosts to visual memory are predicted by theories of visual memory that assume that limits are discrete or continuous, and are consistent with calls to account for perceptual organization and chunking in models of visual memory (Brady & Tenenbaum, 2013). The results reported here confirm that healthy elderly adults likewise benefit from feature redundancy and show comparable fixation patterns to young adults. These data provide additional, novel evidence for the proposal that perceptual grouping eases visual memory limits by linking feature redundancy with reduced pupil size, an indicator of cognitive effort. Though perceptual organization apparently occurs early and rapidly, and does not require devoted attention to be beneficial, it may nonetheless be necessary to invoke general attention processes to fully explain all the ways in which feature redundancy eases encoding and boosts observed memory capacity.

References

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*(4), 362–365. https://doi.org/doi/10.1080/14640746608400055

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109. https://doi.org/10.1037/a0030779

Bredemeier, K., & Simons, D. (2012). Working memory and inattentional blindness. *Psychonomic Bulletin & Review*, *19*(2), 239–244. http://dx.doi.org/10.3758/s13423-011-0204-8

Brown, L. A., & Brockmole, J. R. (2010). The role of attention in binding visual features in working memory: Evidence from cognitive ageing. *The Quarterly Journal of Experimental Psychology*, *63*(10), 2067–2079. https://doi.org/10.1080/17470211003721675

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–84. https://doi.org/doi/abs/10.1111/j.2044-8295.1964.tb00899.x

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Furst, A., & Hitch, G. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*(5), 774–782. https://doi.org/10.3758/BF03198412

Hamilton, C. J., Brown, L. A., & Rossi-Arnaud, C. (2017). Older adults benefit from symmetry, but not semantic availability, in visual working memory. *Frontiers in Psychology*, *8*, 2373. https://doi.org/10.3389/fpsyg.2017.02373

Ishihara, S. (1966). *Tests for colourblindness*. Tokyo, Japan: Kanehara Shuppan.

Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Visual similarity in short-term recall for where and when. *The Quarterly Journal of Experimental Psychology*, *61*(3), 353–360. https://doi.org/10.1080/17470210701634537

Jiang, Y. V., Lee, H. J., Asaad, A., & Remington, R. (2016). Similarity effects in visual working memory. *Psychonomic Bulletin & Review*, *23*(2), 476–482. https://doi.org/10.3758/s13423-015-0905-5

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, *17*(3), 356–372. https://doi.org/10.1080/13506280701766313

Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53*(3), 626–646. https://doi.org/10.1080/027249800410463

Logie, R. H., & Maylor, E. A. (2009). An Internet study of prospective memory across adulthood. *Psychology and Aging*, *24*(3), 767–774. https://doi.org/10.1037/a0015479

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition*, *44*(4), 590–607. https://doi.org/10.3758/s13421-015-0580-9

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: a dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00293

Mathot, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.18

Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, *144*(8), 849–883. https://doi.org/10.1037/bul0000155

Morey, C.C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working

memory. *Archives of Scientific Psychology*, *3*(1), 18–29. https://doi.org/10.1037/arc0000014

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology*, *55*(1), 8–24. https://doi.org/10.1016/j.jmp.2010.08.008

Morey, R. D., & Morey, C. C. (2011). WoMMBAT: A user interface for hierarchical Bayesian estimation of working memory capacity. *Behavior Research Methods*, *43*(4), 1044–1065. https://doi.org/10.3758/s13428-011-0114-8

Morey, R. D., & Rouder, J. N. (2014). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.7). Retrieved from http://cran.r-project.org/web/packages/BayesFactor/index.html

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Park, D. C., & Reuter-Lorenz, P. (2008). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*(6), 1282–1289. https://doi.org/10.3758/s13423-013-0460-x

Quinlan, P. T., & Cohen, D. J. (2012). Grouping and binding in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1432–1438. https://doi.org/10.1037/a0027866

Quinlan, P. T., & Cohen, D. J. (2016). The precategorical nature of visual short-term memory. *Journal of Experimental Psychology-Learning Memory and Cognition*, *42*(11), 1694–1712. https://doi.org/10.1037/xlm0000274

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive Aging and the Compensation Hypothesis. *Current Directions in Psychological Science : A Journal of the American Psychological Society.*, *17*(3), 177–182. https://doi.org/10.1111/j.1467-8721.2008.00570.x

Rossi-Arnaud, C., Pieroni, L., & Baddeley, A. (2006). Symmetry and binding in visuo-spatial working memory. *Neuroscience*, *139*(1), 393–400. https://doi.org/10.1016/j.neuroscience.2005.10.048

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. https://doi.org/10.1073/pnas.0711295105

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, *59*(1), 1–17. https://doi.org/10.1016/j.jml.2008.01.004

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2016). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, 1–10. https://doi.org/10.3758/s13428-016-0741-1

Simons, D., & Levin, D. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267. http://dx.doi.org/10.1016/S1364-6613%2897%2901080-2

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *64*, 532–563. https://doi.org/10.1111/1467-9868.00353

Sun, S. Z., Fidalgo, C., Barense, M. D., Lee, A. C. H., Cant, J. S., & Ferber, S. (2017). Erasing and blurring memories: The differential impact of interference on separate aspects of forgetting. *Journal of Experimental Psychology: General*, *146*(11), 1606–1630. https://doi.org/10.1037/xge0000359

Swanson, H. L. (2017). Verbal and visual-spatial working memory: What develops over a life span? *Developmental Psychology*, *53*(5), 971–995. https://doi.org/10.1037/dev0000291

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *PNAS*

Proceedings of the National Academy of Sciences of the United States of America, *109*(22), 8780–8785. http://dx.doi.org/10.1073/pnas.1117465109

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory*, *13*(3–4), 308–317. https://doi.org/10.1080/09658210344000279

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80–87. https://doi.org/10.3758/BF03196470

WorkWithColor.com :: Color Changes the World. (2013). Retrieved February 13, 2018, from http://www.workwithcolor.com/

Table 1. Frequencies of same and change trials and unique-item and duplicate tests per participant, by set size.

	Set Size				
	1	4	5	6	Total
"Same" Response					
All unique	4	10	10	10	34
Redundancy, test unique	NA	12	12	12	36
Redundancy, test duplicate	NA	12	12	12	36
"Change" Response					
All unique	4	8	8	8	28
Redundancy, test unique	NA	10	12	12	34
Redundancy, test duplicate	NA	8	8	8	24

Note. Trials were divided into two equal blocks.

Table 2. Proportions correct (with standard deviations) by age group, set size, correct response, and colour repetition condition.

Same trials

	Set Size			
	1	4	5	6
Young Group (*N*=21)				
All unique	0.99(.05)	0.75(.19)	0.65(.18)	0.55(.21)
2x, test unique	NA	0.86(.21)	0.72(.23)	0.60(.21)
2x, test duplicate	NA	0.90(.12)	0.75(.19)	0.78(.18)
3x, test unique	NA	0.91(.14)	0.87(.16)	0.74(.18)
3x, test duplicate	NA	0.95(.09)	0.93(.14)	0.85(.22)
Elderly Group (*N*=23)				
All unique	0.95(.17)	0.66(.19)	0.53(.23)	0.51(.25)
2x, test unique	NA	0.75(.24)	0.62(.23)	0.50(.22)
2x, test duplicate	NA	0.83(.19)	0.81(.19)	0.70(.22)
3x, test unique	NA	0.85(.16)	0.72(.26)	0.63(.21)
3x, test duplicate	NA	0.93(.13)	0.74(.22)	0.77(.17)

Change trials

	1	4	5	6
Young Group (*N*=21)				
All unique	0.98(.08)	0.93(.13)	0.88(.13)	0.89(.11)
2x, test unique	NA	0.93(.08)	0.88(.13)	0.85(.19)
2x, test duplicate	NA	0.96(.09)	0.90(.17)	0.85(.20)
3x, test unique	NA	0.99(.05)	0.95(.08)	0.94(.09)
3x, test duplicate	NA	0.95(.10)	0.96(.12)	0.95(.10)
Elderly Group (*N*=23)				
All unique	0.93(.11)	0.86(.11)	0.80(.17)	0.80(.17)
2x, test unique	NA	0.83(.17)	0.86(.13)	0.75(.20)
2x, test duplicate	NA	0.96(.12)	0.80(.23)	0.86(.18)
3x, test unique	NA	0.91(.14)	0.86(.17)	0.80(.21)
3x, test duplicate	NA	0.97(.09)	0.87(.17)	0.89(.15)

Table 3. DIC statistics for each model

Model	number parameters on *k*	DIC
Fine x Test type x Age	**54**	**7328.4**
Fine x Test type + Age	51	7335.0
Fine x Test type	49	7339.7
Gross x Test type x Age	50	7391.1
Gross x Test type + Age	49	7389.9
Gross x Test type	47	7396.7
Fine colour redundancy x Age	50	7375.2
Fine colour redundancy + Age	49	7380.4
Fine colour redundancy	47	7386.5
Gross colour redundancy x Age	48	7446.6
Gross colour redundancy + Age	48	7444.7
Gross colour redundancy	46	7449.9
Test type x Age	48	7444.2
Test type + Age	48	7442.7
Test type	46	7447.9
Age group	46	7586.3
Participant variance only	44	7590.4

Note. The best-fitting model with the lowest DIC is indicated in bold text. Each model was estimated with 25,000 MCMC iterations, with between-participant variance on the *k* parameter in addition to the other factors being compared across models of *k*. The models are ordered from most to least complex. Sections divide models that differ by how the colour redundancy variable was coded. All models included between-participant variance and main effects of age group on the *z* and *g* parameters, plus main effects of colour redundancy on the *g* parameter. *N*=44.

Table 4. Fixation counts and gaze durations toward centre, unique colours, and colour duplicates.

Stimulus Presentation	Mean fixation count	Mean duration fixated
All Unique		
Centre	0.78 (0.31)	280 (134)
Unique	1.51 (0.50)	338 (109)
Duplicates-2x		
Centre	0.83 (0.31)	292 (136)
Unique	0.84 (0.39)	192 (82)
Duplicate	0.58 (0.25)	133 (64)
Duplicates-3x		
Centre	0.84 (0.31)	293 (139)
Unique	0.61 (0.35)	141 (78)
Duplicate	0.78 (0.31)	181 (75)
Retention Interval	**Mean fixation count**	**Mean duration fixated**
All Unique		
Centre	0.39 (0.31)	129 (110)
Unique	2.43 (0.93)	887 (400)
Duplicates-2x		
Centre	0.38 (0.26)	134 (114)
Unique	1.55 (0.65)	588 (308)
Duplicate	0.78 (0.43)	284 (182)
Duplicates-3x		
Centre	0.38 (0.33)	142 (114)
Unique	1.26 (0.71)	522 (335)
Duplicate	0.97 (0.45)	388 (219)

Note. *N*=39. Fixations were trimmed by interest period (stimulus presentation and retention interval). The stimulus presentation interest period included 500 ms prior to the stimulus onset, in order to frequently avoid trimming the initial central fixation. Trials with no recorded fixations to any of these interest areas were excluded from analysis. Standard deviations in parentheses.

Table 5. Average minimum fixation time to the centre, a unique colour, or a duplicate during stimulus presentation in trials with both duplicated and all-unique colours, as a function of age group.

	Centre	Unique	Duplicate
Stimulus presentation			
Elderly group	511 (432)	1164 (473)	1050 (423)
Young group	404 (368)	1188 (438)	1064 (397)

Table 6. Mean pupil sizes (expressed as proportions of baseline) during the retention period by age group, colour redundancy condition, and set size.

	Set Size		
	4	5	6
Young Group (*N*=20)			
All Unique	0.96 (0.08)	0.97 (0.07)	0.98 (0.08)
Duplicates-2x	0.95 (0.07)	0.96 (0.07)	0.97 (0.09)
Duplicates-3x	0.92 (0.06)	0.95 (0.07)	0.97 (0.08)
Elderly Group (*N*=19)			
All Unique	0.97 (0.06)	0.97 (0.07)	0.98 (0.07)
Duplicates-2x	0.97 (0.07)	0.98 (0.07)	0.97 (0.08)
Duplicates-3x	0.96 (0.07)	0.97 (0.07)	0.98 (0.07)

Note. Proportional pupil sizes were calculated by dividing each observation by the participant's average pupil size during the 1000-ms period before the fixation appeared. Standard deviations in parentheses.

Figure 1

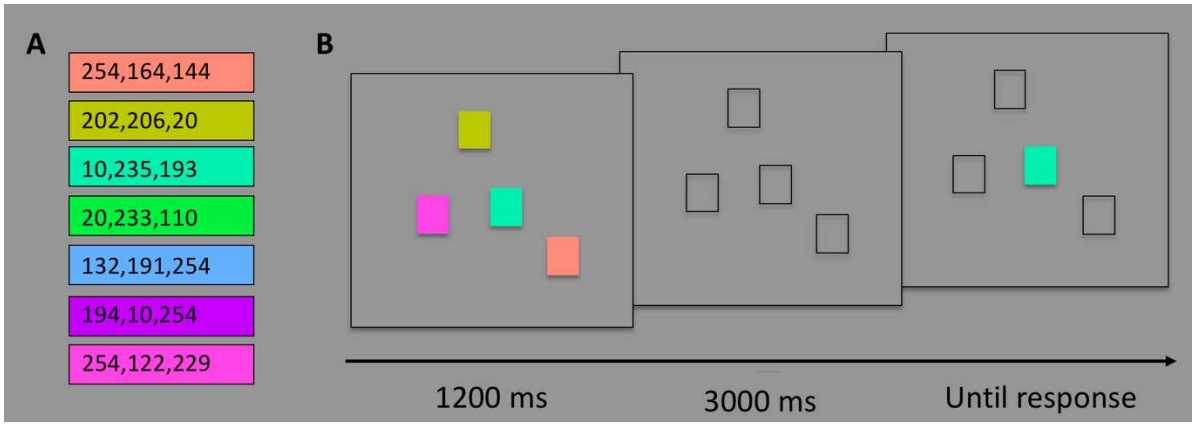

Figure 1. A) Set of seven colours from which stimuli were randomly selected. Note that appearance of colours varies depending on situational variables; these particular shades were selected based on monitor and lighting conditions particular to the laboratory. B) Trial events and their timings. Stimulus presentation was also preceded by a fixation (+), which lasted 400 ms, followed by a 100-ms blank screen. After the participant's response, a 4000-ms delay was imposed before the next trial started. Relationships between square and display size are not to scale.

Figure 2

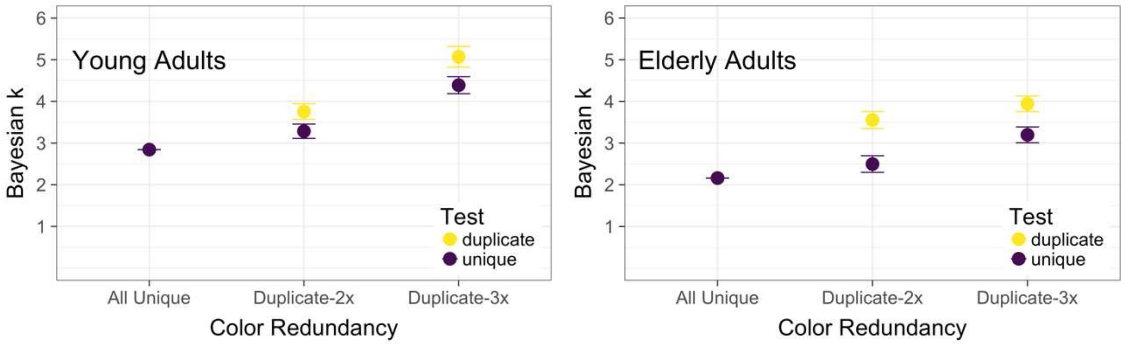

Figure 2. Hierarchical Bayesian *k* estimates. Error bars reflect the posterior standard deviation on the difference between baseline unique colour tests and each other value. *N*=21 young adults, *N*=23 elderly adults.

Figure 3

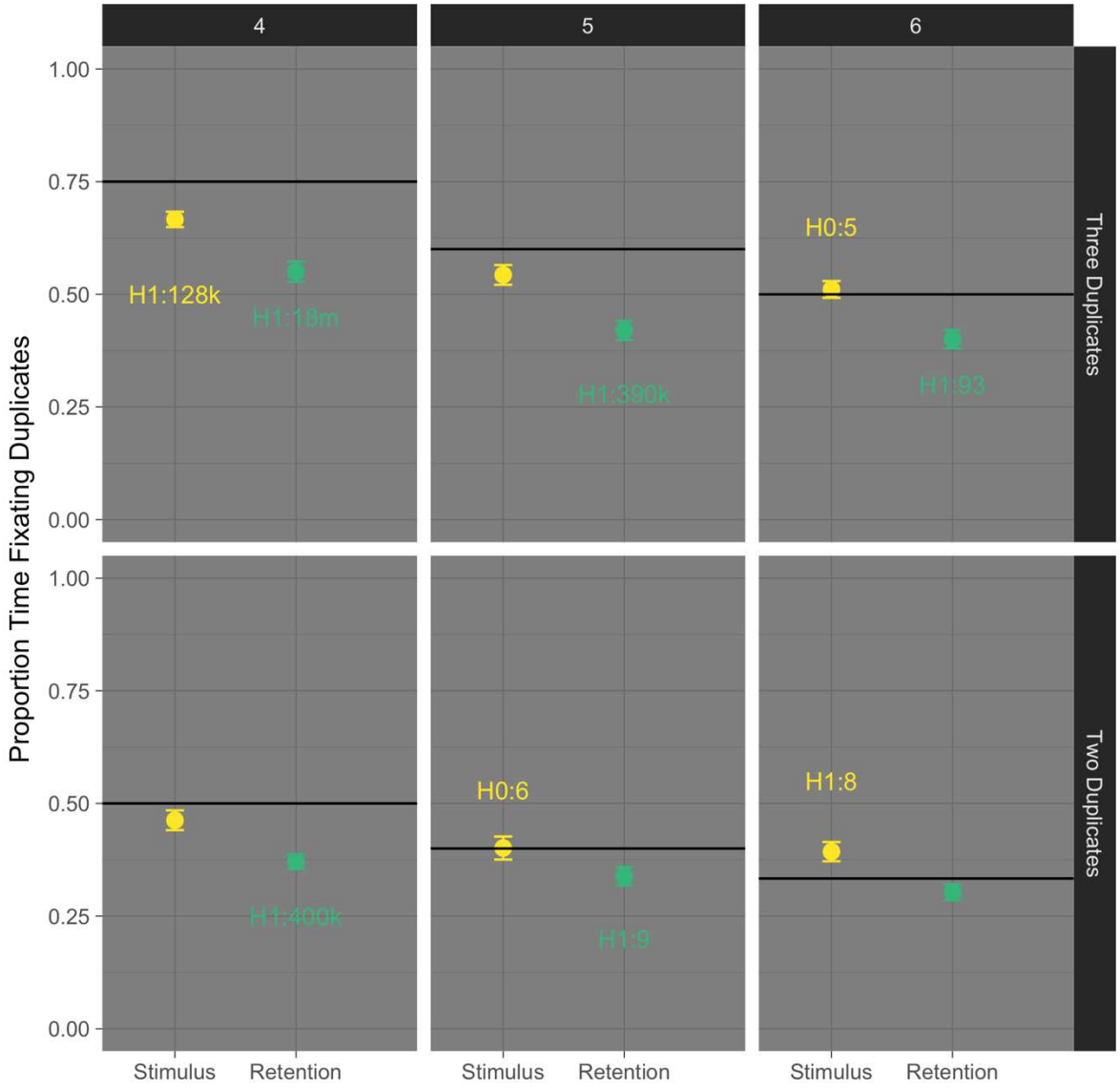

Figure 3. Proportion of time during stimulus presentation and retention period spent looking at a duplicate colour. The black horizontal lines denote the expected value if gaze to duplicates is at chance. Where Bayes factors from independent tests exceeded 3 in either direction, the BF for the comparison between mean proportions and chance are given. H1 indicates that evidence favours the alternative

hypothesis that the mean value differs from chance, while H0 indicates that the null hypothesis is favoured.

Figure 4

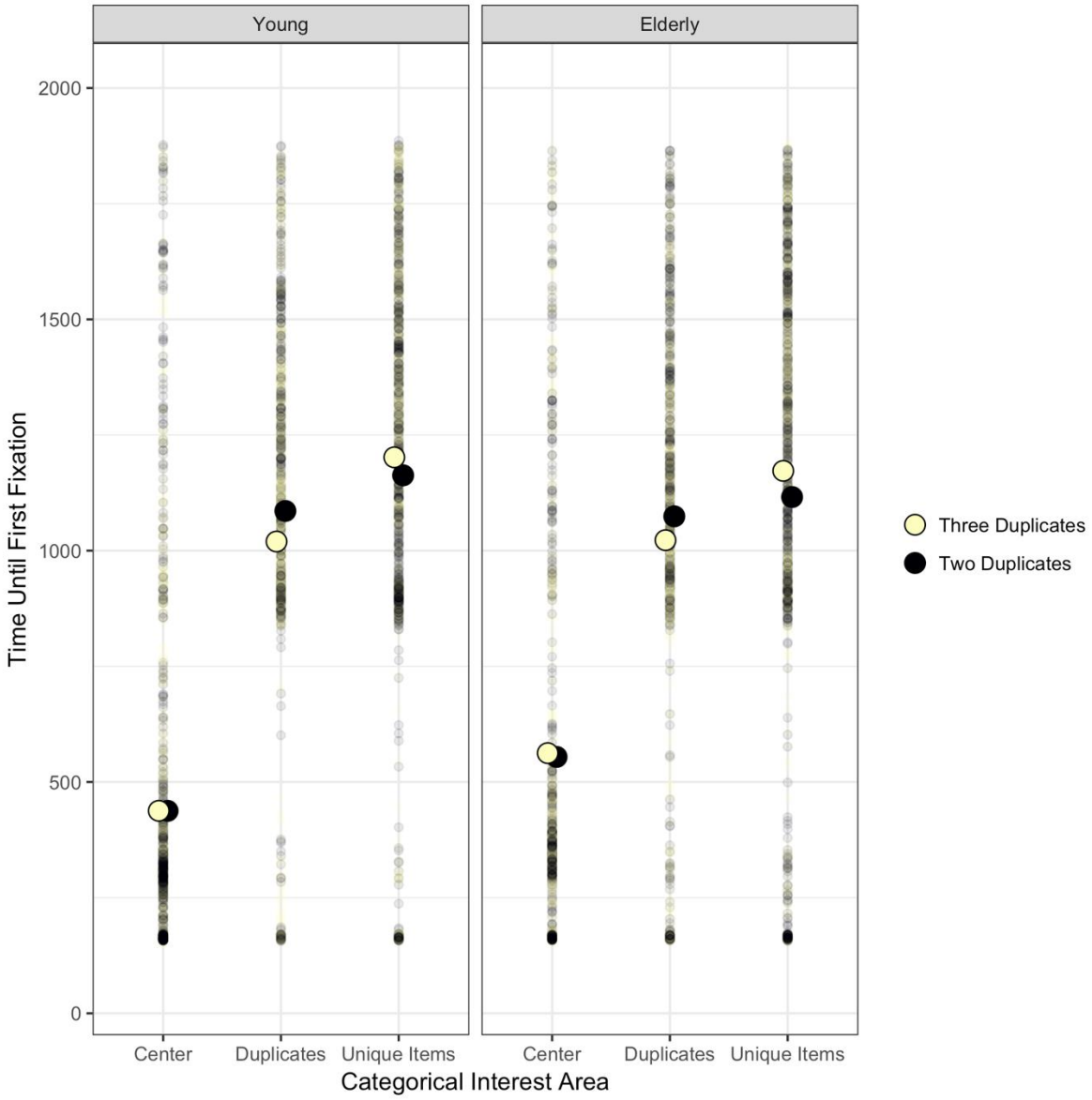

Figure 4. Time (in ms) until the first fixation of the centre, a duplicate colour, or a unique colour, plotted by age group and number of duplicate colours in the array. Bold points represent means, which are overlaid on values of individual data points.

Figure 5

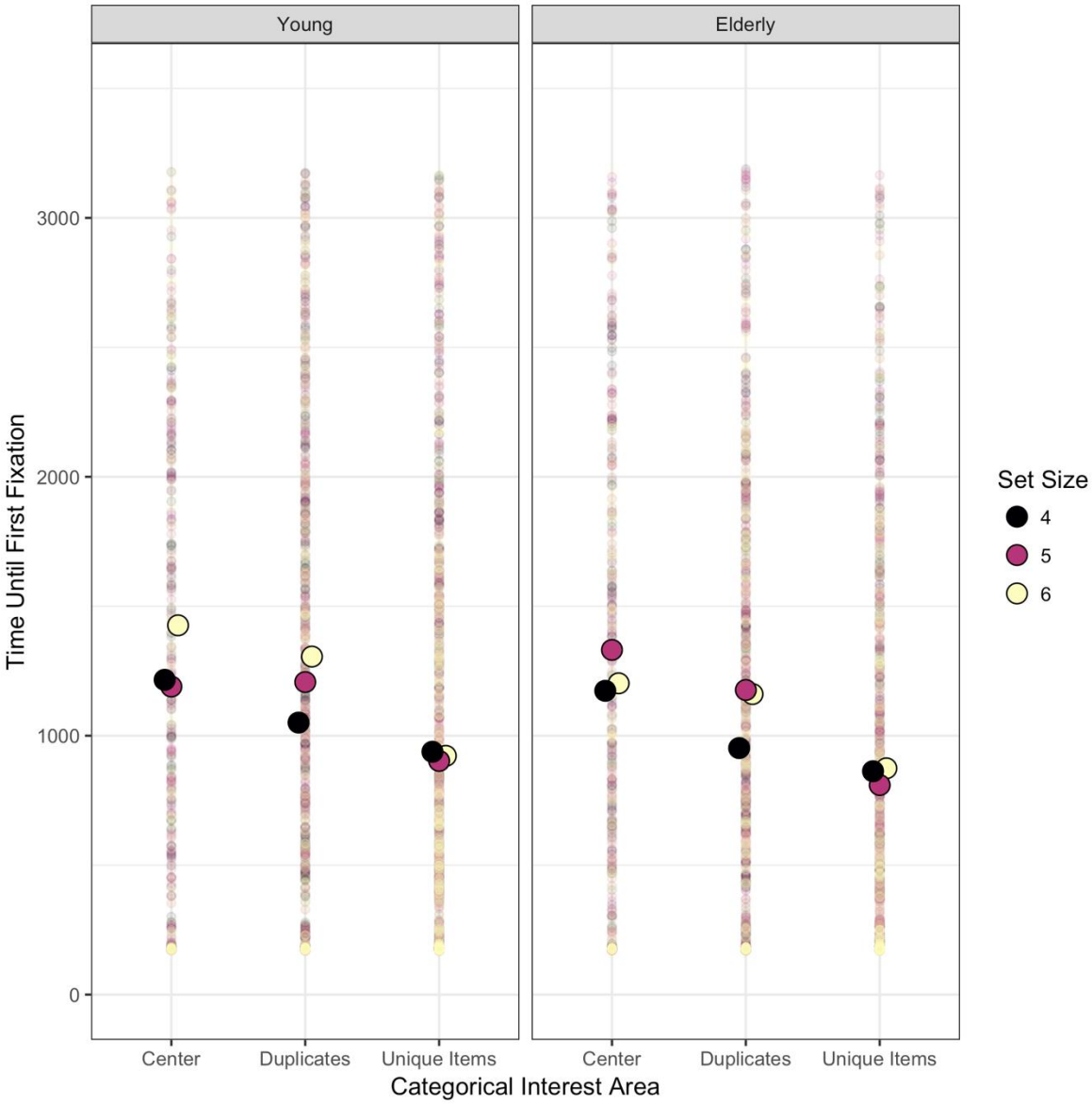

Figure 5. Time (in ms) until the first fixation of the centre, a duplicate colour, or a unique colour, plotted by age group and set size. Bold points represent means, which are overlaid on values of individual data points.

Figure 6

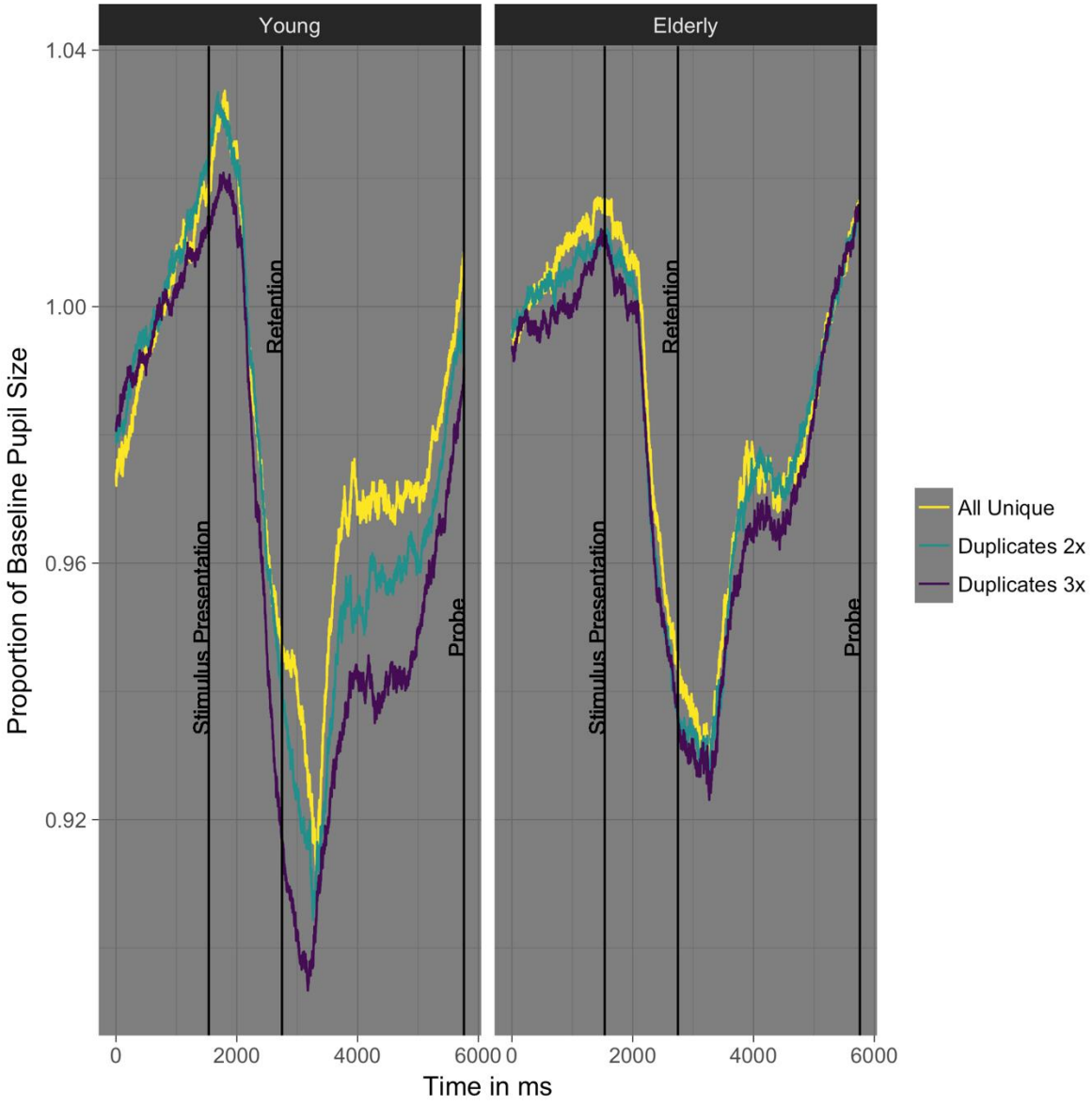

Figure 6. Pupil size (reported as the proportion of an average baseline area measurement taken each trial prior to onset of the stimulus presentation) plotted by trial time in ms, presented as a function of age group and colour redundancy condition. The 0 time marks the last 1.5 seconds of the fixation period;

measurements from the start of the trial acquired before time 0 were used as the baseline.